Title of Project:
Product and Process in TOEFL iBT Independent and Integrated Writing Tasks: A Validation Study

Researcher:
Liang Guo
Georgia State University
guoguo18@gmail.com

Research Supervisor:
Dr. Sara C. Weigle

Project Summary:

This study focuses on comparing text based integrated writing tasks (writing based on source materials) with more traditional independent writing tasks. Despite their many potential benefits, there have been relatively few studies of integrated writing tasks in the literature of second language (L2) writing assessment, especially when compared with the abundance of research on independent writing tasks. The very limited studies that have examined integrated writing tasks have focused either on the product or the process of the writing performance to validate such tasks used in a testing context. To clarify the construct inherent in text based integrated writing and to verify the previous statements that have been made about the task as a promising task type, it is necessary to conduct quantitative textual analysis of the essays composed by test takers and obtain qualitative information on the writing processes as well (Bachman, 2004). The central issue addressed in the current study, therefore, is concerned with whether writing products and processes elicited by TOEFL iBT integrated and independent writing tasks vary with task type, perceived writing quality, and the academic experiences of test takers in accordance with theoretical expectations.

In order to address the issue of the influence of task type, writing quality, and test taker experiences on writing products and processes, quantitative analyses were conducted on the integrated and independent essays. In addition, qualitative methods were used to explore the writing processes generated by the two tasks respectively. The quantitative section focused on whether textual features vary along with the task type, the perceived writing quality, and the academic experiences of test takers. The qualitative section investigated how writing processes are related to the same three factors. In the following paragraphs, I explain how the data were collected and analyzed, provide a summary of the findings, and suggest possible implications.

To answer the question whether the textual features vary along with the task type, the perceived writing quality, and the academic experience of test takers, 480 TOEFL iBT essays (from fall 2007) provided by Educational Testing Service (ETS) were collected. Two hundred forty test takers from different linguistic backgrounds responded to an integrated task and then an
independent writing task. For the integrated writing task, the test takers read a passage and a listening passage which addressed the same topic but offered a different perspective from the reading material. Then, the test takers were required to summarize the listening passage and state how it challenged the points made in the reading passage. For the independent writing task, the test takers were required to write an argumentative essay on a given topic, where they supported an opinion on a given topic using their prior knowledge and/or experiences. Co-Metrix, a computational tool, was used to analyze textual features of the collected essays.

To explore whether writing processes change with the task type, the perceived writing quality, and the academic experience of test takers, 20 English as a second language (ESL) writers were recruited to participate in think aloud sessions to articulate their thoughts while composing. They were given the same tasks as in the quantitative section. Prior to conducting the think aloud writing tasks, training based on Ericsson and Simons (1993) was provided to help the participants to become familiar with the think aloud protocols. Think aloud data were fully transcribed, and writing behaviors were then identified and analyzed by the researcher and re-examined by another experienced researcher.

To investigate whether textual features vary along with the task type, discriminant functional analysis was conducted and the results suggest that linguistic features, mainly lexical features, can predict essay membership with 100% accuracy. The study found that as compared to the independent essays, the integrated writing was characterized by the more frequent use of verbs in 3^{rd} person singular present tense and a larger number of modifiers per noun phrase, which indicated a more detached way of writing and an informational prose style (Biber, 1988). At the lexical levels, the integrated writing was also marked by heavy use of concrete words and meaningful words. This finding might be an artifact of integrated writing as the writing content was highly controlled. The fact that many more concrete words were used also suggested that the integrated writing tends to be more context-independent than the independent writing, another characteristic of formal, academic writing (Crossley & McNamara, 2009).

To study whether textual features vary with the perceived writing quality (as determined by the essay scores) within each task type, a regression analysis was conducted using the scores as the dependent variable and Coh-Metrix indices as the independent variables. The results illustrated that the linguistic features can at least partially predict evaluation of writing quality for either the integrated or the independent essays.

Comparing the predictor indices of the integrated essays with those of the independent essays, it can be seen that there are both similarities and differences. Both types of essay scores were found to be positively correlated with text length and lexical sophistication and negatively correlated with verbs in base forms. The results indicated that the longer essays and the essays containing sophisticated words tended to be rated higher in both tasks regardless of whether these items were specified in the scoring rubric. Verbs in base forms, through further investigation, were found to be mainly grammatical mistakes (without the necessary morphemes to indicate tense, number or person). The negative correlation indicated that the essays demonstrating higher grammatical accuracy were more likely to be scored higher. The higher
rated integrated essays had a significantly higher score in semantic similarity, which represents conceptual similarities among sentences. No cohesive devices that could predict the essay scores in the independent writing tasks were identified. Verbs in the past participle form were a significant predictor of the integrated essay scores but not for the independent scores.

The last quantitative question focuses on whether the linguistic features vary with the academic experience of the test takers within each task type. Writers with more experience at the tertiary level should gain more familiarity with academic writing because of the exposure and practice they have in understanding and communicating in such activities (Kutoba, 1998). The study, however, found that the majority of the linguistic features did not significantly discriminate the two groups of writers. Academic experience at the tertiary level does not seem to leave a noticeable trace in the linguistic choices made by the writers while constructing the integrated and independent essays. Furthermore, the very few features that showed a significant difference across the two groups of test takers do not lead to a score difference for either of the tasks.

The qualitative section first looked at the relationship between the writing behaviors and the task type. The results of a t-test illustrated that the two tasks—the integrated and the independent—provoked similar behaviors in the L2 writers. Wilcoxon Signed-rank test results showed that the participants did not differ significantly across the two tasks in “verbalizing one’s writing,” “planning and rehearsal,” “commenting on one’s writing product,” “commenting on one’s writing process,” and “analyzing the task.” However, the integrated writing task generated significantly lower percentage than the independent writing task in “reading the instruction,” “global planning,” “reading one’s writing,” and “revising and editing.” Due to the different nature of writing across the two tasks, the two tasks also produced writing behaviors that were unique to each of them. The integrated writing task had “summarizing source texts,” “referring to source texts,” “commenting on understanding of source texts,” and “commenting on relationship between source texts.” Meanwhile, the independent writing task had a particular category of “positioning self,” which indicated that the test takers spent time evaluating which side of the argument they chose.

As for whether the writing behaviors varied with the writing scores, two experienced raters using the rubrics provided by ETS scored the essays. The participants were divided into high and low performance groups for the integrated and for the independent tasks. For the integrated task, Mann Whitney test revealed no significant differences in writing behaviors across the two performance groups. A similar finding was reported for the independent task except for the category of “revising and editing.” Namely, the high performance participants revised and edited significantly more frequently than their counterparts.

The very last qualitative question looked at the writing behaviors in relation to the academic experience of the participants within each task. The study found that the integrated writing task did not generate significant differences in terms of the types and the frequency of the writing behaviors. As for the independent writing task, the two groups of participants did not differ significantly from each other in the number and type of writing behaviors. However, in terms of
frequency, Mann Whitney test revealed that the undergraduate participants produced significantly more “analyzing the task” behavior than the graduate.

This study aimed to clarify the issue whether writing products and writing processes elicited by TOEFL iBT integrated and independent writing tasks vary with task type, perceived writing quality, and academic experience of test takers in accordance with theoretical expectations. The study found that, on one hand, the textual features varied with the task type and the perceived writing quality. Certain textual features were reported to be able to predict essay membership with 100% accuracy. For each task type, textual features were also found to be able to predict the essay scores (the perceived writing quality). However, as for how textual features are related to the academic writing experience of the test takers, no significant differences were identified within each of the tasks. On the other hand, the majority of the writing behaviors were not found to vary with the task type, the perceived writing quality, or the academic experiences of the test takers.

The findings have significant implications for both L2 writing assessment and L2 writing instruction. The rich empirical evidence revealed that the integrated writing and the independent writing tasks did elicit different writing performances and, thus, affirmed the proposed rationale for the combined use of the two tasks, namely broadening “representation of the domain of academic writing on the test” (Huff et al., 2008). Therefore, the question whether and why the two test items should be used simultaneously in assessing academic writing is answered. The study also investigated whether test performance varied with the writing quality perceived and the academic experience of the test takers. The results not only helped to clarify the score meaning in each of the writing tasks but also to validate the scoring rubrics used. Taken together, all this information helps to clarify the link between the observed score and the underlying writing ability being assessed (Chapelle, Enright, & Jamieson, 2008), thus building a more comprehensive picture of L2 writing assessment, especially in regards to integrated writing tasks.

As for L2 writing instruction, the differences identified between the products and processes across the two tasks suggest that these two types of writing represent at least two different aspects of academic writing ability. Instruction in the more conventional independent argumentative writing by itself might not suffice and fully prepare L2 writers to cope with academic writing tasks. Writing instruction and learning, therefore, should include sourcing texts and the synthesizing of these texts into writing to provide L2 writers with adequate exposure to such writing activities and to develop the corresponding writing ability that is integral to academic activities of higher education (Cohen, 1998; Hirvela, 2004; Bachman, 2004).
References


Cumming, A., Kantor, R. Baba, K., Erdoosy, U., Eouanzou, K., & James, M. (2005). Differences in written discourse in writing-only and reading-to-write prototype tasks for next generation TOEFL. Assessing Writing, 10(1), 5-43.


McNamara, D. S., & Graesser, A. C. (in press). In P. M. McCarthy & C. Boonthum (Eds.), Applied natural language processing: Identification, investigation, and resolution. Hershey, PA: IGI Global


